MR: Mike Rabinovici here and this is Dimodelo's "Conversations with Data Warehouse Experts" podcast series. Our mission in these podcasts is to speak to the best minds in the data warehouse and BI space, and to get their take on the state of the business and find out what they think the future holds. We are also committed to do it in a way that adds value to the tech profession, while at the same time is clear and compelling for the business managers and executives they collaborate with every day.

This podcast is brought to you by 7W Data, your go-to portal for all things data. Subscribe free for the latest news and events at www.7Wdata.com. Follow on Twitter @7Wdata.

In today's episode, it is my pleasure to speak to John Johansen, Senior Vice President, Consulting Services Strategy & Innovation at Majesco, a leading provider of insurance solutions and services. He's a highly skilled consultant and data architect known for his ability to help insurers build effective data warehouse organizations. He's also a prolific writer who regularly publishes thought leadership pieces on the business of data and insurance. Welcome John!

JJ: Well, thank you Mike, it's good to be here.

MR: I'd like to begin by asking you to tell us a little bit about your background, where you started, and how you arrived at Majesco.

JJ: Well, if I'm going all the way back, I actually started writing dBase II programs for my parents' business when I was 13 years old, and I've been a data guy ever since. Back in 1997, I started Agile Technologies with two partners, basically organized around helping insurance companies to make more money using technology. We built that business for about 17 years, to about 100 folks, and sold it to Majesco—a publicly-traded global software and services company—back in 2015. Ever since, we've been growing our data teams and delivering larger and larger projects to our insurance customers.

MR: Where have you seen the most significant changes in the data warehouse and BI space over the last, say, five years as it relates to the insurance industry?

JJ: In the insurance space, we've seen three big changes. The first is that data has moved to people that are not just actuaries. So insurance companies have had a blessing—they've had data scientists on their teams for decades in the form of their actuaries. But what we're seeing is much broader access to data warehousing and data assets, and more use cases outside of those traditional lanes.

One of the other things we're seeing is insurers recognizing the real business benefit of pervasive modeling. A few years ago, property insurers began to model every catastrophe-exposed risk using very specific modeling software. Up to that point, they'd really just modeled portfolio or exceptions on a portfolio basis, but here they wanted to model every single risk in their portfolio. It was a real game changer for them in terms of the efficiency of capital, knowing where the real risk was and making sure they were getting paid for that risk. And insurers are bringing that idea of "let's model every risk" now to every line of business. They're doing that with their business owners' policies, with home owners'

policies, and they're basically saying: how do I build a model that will help me quantify the real risk on these risks.

Finally, we're seeing our clients recognizing that they can compete with data and that they really need to get their usage of data correct. While it's an older study, about four years ago IBM and MIT put together a paper and they found that people who were using data well were getting better at it faster than the people that weren't using data well. What that really means is that the people who are good at it are getting better at it, and they're extending their lead. I really think that over the last 5 years our clients have been embracing that reality and realizing that the need to get great at using data to advance their businesses.

MR: That's a great point because I speak to our clients when they look at these types of opportunities. Do you find that insurers are starting to appreciate that accurate, timely BI is no longer a luxury, but a necessity to continue to survive and thrive in the business?

JJ: Absolutely. One of the things that's happening is the alumni network of certain organizations is kind of moving around the insurance industry, and they all kind of look around and say: how can I be expected to run a business in an environment where I don't have access to great BI? They're saying: at my last place, I was able to get this and it was easy because my last company was a great provider of data and BI. And now they're saying: I need this for my business or else I'm flying blind. I need the instrumentation to be such a way that I can make the same types of business decisions that I was able to before.

MR: From your vantage point, what are the key pain points in the insurance business when it comes to data warehousing and ETL?

JJ: We see that there's very real business currents that present data warehousing and ETL challenges to insurers. Some just have to do with the way the insurance business works. The supply chain for insurance includes independent captive agents, specialized agents, third-party claims processors, brokers, carriers, reinsurers, and they're all trading information, and some of them are still using Excel. So that creates some real challenges in data warehousing and ETL.

Another area is that insurers seem to be in this never-ending M&A cycle. Years ago, our first customer got bought by a company called Ace, and more recently Ace bought a company we've probably all heard of, called Chubb, and it seems like it's the circle of life in the insurance industry—that M&A is one of these things that's going to consistently go on. That, again, creates a real integration challenge which really manifests itself in data.

Insurers aren't immune to the pressure of new entrants either. So there are companies like Lemonade, which is a fully digital insurance company that's aimed directly at the buying patterns of millennials; a company called Metromile, which charges you on your auto insurance based on how many miles you

drive… that pressure is very, very real, let alone the threat of a competitor like Amazon deciding they want to sell insurance and price it based on your purchase patterns, or something like that.

MR: I think it's just a matter of time, based on how they've been growing. Do you think that current data warehousing tools are adequately addressing these pain points?

JJ: Well, if I'm being honest, the existing custom code-based ETL tools tend to be part of the problem. I've seen clients with 65 similar, but not exactly the same, data feeds, and they implement those data feeds and those old school tools in—I don't know—130, 195, 290 distinct ETL squares. Those tools make it really easy to clone your way into a real mess, and before you know it, you kind of have a support nightmare.

That said, there are tools in the market that are challenging the ETL establishment. Certainly, your tools, the tools from Wherescape, are making a big impact. I also think that some of the data wrangling tools are beginning to emerge into tools that will allow for what we call zero-ETL data prep. Those will make a big impact as well.

MR: Speaking of the big data warehouses themselves, from your vantage point, some pundits have over the last few years heralded the death of the data warehouse. What's your take on that from the insurance business perspective?

JJ: I think the rumors of the death of the data warehouse have been greatly exaggerated ☺, to mangle a cliché. We're in a golden age of data. Nobody wants to pay for something that feels like it's 15 years old. So, while it may not show up in the marketing literature, I think the data warehouse as a functional idea is alive and well. Our clients are telling us that there will always be room in their architectures for a production-class data store that presents consistent, trusted data to business users across the enterprise. Even when our clients are employed data lakes, there's always a conformed set of structured user information that, when taken together, look a lot like a data warehouse. We're also seeing the rise of less structured, less governed data sets used for exploration and discovery. And these exploration platforms are growing in addition to the data warehouse, certainly not in place of the data warehouse.

MR: Let's switch gears for a minute and talk about the cloud for a moment because the discussion about it has been so pervasive. What has been the impact on data warehousing projects in the insurance industry as more and more companies are adopting the cloud?

JJ: Insurers have moved pretty carefully into the public cloud for obvious reasons. No one wants to hear that their health information has been leaked because some upstart cloud vendor didn't apply a patch, and the network got compromised. This kind of concern has been a real inhibitor to public cloud-only insurer models. And when it comes to data warehousing, our global insurance clients have real data sovereignty issues. Insurance is regulated country-by-country, and the information that can and cannot be collected differs by country. Some regulators have said that personal information can't be stored outside their country borders at all. Others have ruled that they'll apply the same stringent compliance

regulations to all the data stored within their country or borders, even if the insurer doesn't [??] citizens (@10:20).

So cloud providers have, for the most part, created solutions to these challenges, but they're fairly recent developments. Again, that's kind of slowed the adoption rates of those kind of public cloud implementations. But at the end of the day, insurers have a lot of data. And even private cloud data warehousing solutions end up being limited by bandwidth. It's actually, I think, pretty telling that cloud data warehouse vendors out there all have some kind of branded initial load that uses what I like to call "Fedex bandwidth" as part of their process. Those initial loads get put on some type of medium and shipped via Fedex in order to deal with those initial loads. Because a lot of times those initial loads are trying to put a whole lot of bits down what are still relatively small pipes.

But net-net, I think we're seeing insurers adopting cloud models because of their scale and the agility benefits they bring, but they're pretty cautious about some of these important details.

MR: What's your take on the role automation can play in making the design and build of these data warehouses faster and more cost effective, especially given the current increase in data flows and complexity?

JJ: We've talked about how ETL is a big problem we're all trying to solve. I personally think that's a big piece of the value equation driving data lakes. But automation is going to shorten our cycle time, and it really should allow us to demonstrate agility in our data projects, where we can drive multiple fast iterations that will deliver valuable data to the software development that's going on on the front end. And I think the market's ready for another dynamic here too, and that is this whole concept of designing for agility and designing for automation. We see concepts like data lake and data [?? - @12:10] really driven by the need to have the backend data processes better support agile development models. So people are willing to trade some design purity for the benefit of better support for agile development and really, more importantly, trade it for organizational agility, which is something almost all our customers are striving for today.

MR: When you and I spoke a few days ago, when we were prepping for this podcast, one of the things that we were on the same wavelength about was the need to educate the business suite a little bit more about what's going on on the IT side of the business, because that's where we see a lot of dichotomy. We find the IT folks pretty knowledgeable and the knowledge gap comes when we talk to the mid-level managers and to C-suite. If you were to recommend a resource—be it a blog, a video, a book—for the executive side of the business to be able to learn a little more about what their IT folks are trying to do on the BI and data warehousing end, what would it be?

JJ: I really want to amplify the important point you're on here. Back in the 80s, every great manager, every great executive, needed to be a great project manager. All of a sudden, they all needed to learn about Gantt charts, and making sure that timelines were represented properly. What we're now really seeing is that, today, everybody needs to be really great at data and finding resources to help advance

those executive teams is really an important step in the process that the industry's going through right now.

Fortunately, there is a lot of content out there for business managers to use. There's TED Talks, which are nice little bite-sized chunks of information. There's one in particular that I like to recommend to business folks from David McCandless, and it was about data visualization and the role that data can play in an organization. He does a really good job presenting some interesting visualizations and also really quantifying for people how data can play an important role in the organization.

The Harvard Business Review also has an excellent series of articles by Tom Davenport. They do a great job teeing up data solutions for business leaders and, as everyone knows, the Harvard Business Review does a really good job of including that in business terms. They never really lose sight of the bottom line, which is something that kind of jumps out at us.

Finally, what we are seeing too is that universities have jumped on providing non-degree programs that are aimed at executives and managers. These can be as simple as week-long certificate programs, but they all provide a deeper dive for business people. And I'm on the advisory board of a week-long big data program at Rutgers University in New Jersey, so there's a shameless plug we can also include! If you surf to data.rutgers.edu, you can find out more details on that.

MR: Thank you. We'll definitely include that information in the notes to our podcast.

In September, you wrote a really interesting article on "Underwriting Data and Lessons from the Professional Golf Association". To be honest, you just had me at "golf", so I read it, but can you share the highlights of the article with our listeners?

JJ: That was one of the pieces I wrote that was a little more fun to write than some of the others. It was a pretty simple premise. I participate in golf pools for each of the four majors in golf every year with some folks I play golf with at my golf club. This year, I decided to have big data help me make my picks for each of these four tournaments. What you do is, you pick a portfolio of golfers and then whoever picks the best portfolio of golfers wins. CBS Sports runs a Monte Carlo simulation of 10,000 rounds of golf, and then ranks the predictive and performance of each of the golfers based on those simulations. I picked my players for two events using this approach. I'm happy to spoil the story, Mike, if you want me to…

MR: Please.

JJ: Well, let's just say that big data picked two very, very good teams of golfers and it was a real victory for big data over Las Vegas odds makers.

MR: That's great. I love that. Now, I want to switch gears and have some non-tech fun, just so our audience can get to know our guests a little bit better. So you ready for some questions?

JJ: Oh, my gosh! Sure. Let's see where this goes.

MR: Alright. So, #1: what is your current go-to TV show that you're either watching or binging on?

JJ: Well, I'm not sure if this fits the non-tech fun, because my current binge is a series on AMC called *Halt and Catch Fire*.

MR: Oh my God. It's my favorite show! I can't believe they're ending it this season!

JJ: I have to warn you, I actually saw an article on the end of the series a little while ago, so I'm only about midway through season 3. They've just kind of begun to [?? @17:28] the Internet… I don't know if that Internet thing is going to catch on or not, but it's been…

MR: [laughs] I guess we'll have to wait and find out. I thought it was a brilliant show, well written, great characters. Highly recommended. Next question: What is the best non-tech book you've read in the past few months that you'd recommend to all your friends?

JJ: I'm a huge reader, so I end up being a sucker for some of the books that come across the hype cycle. One I picked up fairly recently is Dan Brown's new book, *Origin*. True to the *Da Vinci Code*, this is a really enjoyable story full of a lot of historical facts, but also a page-turning kind of fiction story just like he usually delivers. So when friends ask me what I'm reading, and they're not my geek friends, I usually recommend *Origin*.

MR: I will add it to my reading list, especially based on your picking of *Halt and Catch Fire*. Last question, John. Where can our audience find out more about you and Majesco on the interwebs?

JJ: Well, I publish as you mentioned a lot of thought leadership papers that I write at majesco.com. People call follow me on Twitter @agilejj. And I am, of course, on LinkedIn at John Johansen. And let's not forget my link to bigdata.rutgers.edu, right?

MR: It was a pleasure speaking with you, John. Thanks so much for your time.

JJ: Thank you for the time. It was a pleasure.

MR: This was another podcast from our Conversations with Data Warehouse Experts. Til next time!